DOCUMENT RESUME

ED 129 739 SP 010 452

AUTHOR Carlson, Robert V.; Park, Richard
TITLE Teacher Evaluation: Relevant Concepts and Related Procedures.
PUB DATE Jun 76
NOTE 33p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Data Collection; Effective Teaching; *Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; *Models; Student Evaluation of Teacher Performance; Teacher Behavior; Teacher Characteristics; Teacher Education; *Teacher Evaluation; Teacher Improvement; Teacher Qualifications; Teaching Quality

ABSTRACT

The present purposes of teacher evaluation commonly include: (1) professional growth for improvement of instruction; (2) clarifying goals and objectives; (3) measuring progress toward those goals; (4) clarifying inservice needs; (5) judging the contribution of the teacher to pupil progress; (6) determining salary; and (7) determining employment status. Three conceptual frameworks of teacher evaluation are discussed: appraisal based on mutually derived objectives; appraisal based on student learnings; and appraisal based on teacher behavior. The question is which approach or combination of approaches should be used. Several factors are discussed which must be considered in designing a "new" teacher evaluation system: baseline assessment; definition of a decision-making process; clarification of purposes of the evaluation system; identification of alternative approaches; identification of type of data needed; information sources; data collection methods; management structures; selection of plausible design features. Various implementation stages for a teacher evaluation system are outlined, and a model for evaluating the new evaluation system is presented. (JMF)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED129739

TEACHER EVALUATION: RELEVANT CONCEPTS AND RELATED PROCEDURES

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: SP EA
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. TM

Prepared By

Robert V. Carlson
Professor
Organizational and Human Resource Development Program

and

Richard Park
M.Ed. Graduate
Organizational and Human Resource Development Program
College of Education and Social Services
University of Vermont

Revised June, 1976

SP010 152

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FOREWORD

The document which follows presents a practical review of emerging concepts concerning teacher evaluation. It also specifies some practical procedures in moving from an existing system for teacher evaluation to a newer system. Specific questions, and steps and issues, are presented which should permit the practitioner a helpful guideline in embarking on such a journey.

A deliberate attempt has been made to minimize jargon and to present numerous alternatives, thus maximizing the potential options which present themselves in designing a "new" system.

An attempt has been made to be comprehensive but succinct. We consider what follows as a starting point and not an end. Helpful feedback and suggestions are urged.

RVC
RWP

I. FANTASIZED ETIOLOGY

Teachers have evaluated students since before there were formal schools (i.e., "She/he's bright" or "she/he picks things up so quickly"). Students have evaluated teachers for just as long (i.e., "She/he's an old hag" or "She'he's neat!").

Evaluation of students probably first become formal for promotion or selection to a higher rank or training. Perhaps the observation that this evaluation improved performance increased its use, even when promotion was more or less automatic. Evaluation of students has taken the form of trait descriptions--bright/lazy, pass or fail, a letter grade (A-D, or C,S,U,I), a number between 0 and 100 (representing a percentage mastery), a normative statistic (stanines, etc.), teacher/student conferences—or self-evaluation of the learner consensually validated by the teacher. A rather ingenious system for inter-correlating these systems has arisen:

97% = bright = A = pass
58% = lazy or stupid = F = fail

Formal evaluation of teachers is probably as old as that of students though it has not achieved a high profile, usage, or popularity. Evaluation of teachers has taken the form of character traits (cruel, pansy, fine-mannered, diligent), behavior outside the school (socialist, alcoholic, avid church goer), student/teacher interactions within the classroom (she/he can't keep order; she/he is too severe), and knowledge of subject area (heretic, authority, bluffer).

One might wonder why student evaluation is so popular and sophisticated whereas teacher evaluation is still relatively uncommon and crude. In a competitive society such as ours, being evaluated is threatening (to some degree) to anyone lacking an extraordinary self-directedness and self-worth. It is not just negative feedback that is frightening; genuine positive feedback is equally difficult to receive. But receiving is only the half of it (if it were the whole of it, we would be evaluating like crazy). It is also scary to give genuine, direct evaluations of a positive or a negative kind to others. This seems to have the most veracity when the people involved are nearer to being peers without being felt to be true peers (i.e., equals). Thus, it is easier for a principal to evaluate a student teacher than a senior staff member. It is easier for a teacher to evaluate a student than a student teacher. It is easier for a student to evaluate a teacher than for a department chairperson to evaluate a teacher. To get around this threatening aspect of evaluation, some systems have been designed to use anonymous feedback thus reducing the threat level for the evaluator and perhaps the evaluatee. Not surprisingly, it tends to improve the quality of the information. Open evaluations tend to produce a "halo" (everyone is good) or a "norming" (everyone is about the same) effect.

This simplistic formula excludes self-evaluation which, though a nearly continuous process, is not often conscious, purposeful, structured, or formalized. The threat level of self-evaluation is in part determined by self-concept, past successes of self-evaluation, and commitment to change.

Grossly stated, teacher evaluation has been difficult and infrequent because: a) schools are not authoritarian enough to greatly discriminate between the status of teachers and the department chairperson or principal (the typical evaluators); b) schools are not equalitarian or non-competitive enough to permit non-threatening evaluation by other equals; c) teacher do not have a high enough professional commitment to change, or self-assuredness to stimulate self-evaluation; and/or d) educators have suspicions concerning the validity and reliability of teacher evaluation instruments and processes.

II. PRESENT PURPOSES OF TEACHER EVALUATION

A question which precedes "evaluation" per se is whether one wants to assess or evaluate. Assessment involves merely the measurement of an input, process, or outcome. Evaluation, however, involves making a judgment concerning that input, process, or outcome.

If the purpose is merely to know, then assessment is in order. If the purpose is to maintain, change, increase, or decrease a behavior, the route to take is evaluation.

One could choose to simply assess the number of second grade teachers using basal readers in their classroom, without making a judgment as to whether this was desirable or undesirable. Or, one could choose to assess the amount of time the average sophomore spends in the school's library.

Evaluation, on the other hand, implies that a judgment, formed through emperical research, values, reasoning or feelings, is being made as to the desirability of basal readers in second grade, or time logged in the library by sophomores.

If one decides she/he wants to evaluate, not merely assess teacher performance, the present purposes for teacher evaluation commonly include:

1. professional growth for improvement of instruction
2. clarifying the goals and objectives of a department, building, or district
3. measuring progress toward those goals and objectives
4. clarifying in-service needs of a department, building or district
5. judging the contribution of the teacher to pupil progress
6. determining salary
7. determining employment status.

Evaluation to promote professional development entails an instructor getting feedback from students, peers, supervisors, or test outcomes to enable that person to define their needs (the discrepancy between how they and others would like them to be, and the way they are) for change of that person's behavior. Just as importantly, it tells them what they are doing well, what they don't need to change, and what they might help their peers with. Evaluation should not expect that all teachers can be good at all things. Unrecognized or ignored weaknesses are destructive. Recognized weaknesses that are dealt with by remediation or capitalizing on strengths need not be detrimental to learning. Team teaching or careful matching of students with teachers can more than compensate for weaknesses in teachers. A teacher who has difficulty being a disciplinarian need not be given students who have a high need for a person to continually set limits for them.

Because evaluation is a judgment formulated from the congruence or discrepancy between expectations and actualities, formalizing the evaluation process helps to surface those sometimes hidden expectations (desires, goals, and enabling objectives) both of programs and people. Dissatisfaction with an educational system may be due to one party (i.e. parents) not understanding the instructional objectives (i.e. affective domain) of teachers. Or teachers may not realize some objectives they were expected to accomplish with their youngsters (i.e. developing career skills). It may also point out that appropriate resources (i.e. time and money) are not being channeled toward the most important objectives.

Use of the evaluation system designed after these clarified expectations surfaces the progress made from the actual to the ideal. This information then becomes feedback to help redefine or reaffirm needs and appropriate objectives and activities. Continued discrepancies between desired objectives and activities and what is really happening may point toward the need for additional resources: materials, time, in-service training, additional personnel. Or, people may decide to change their expectations.

If evaluation focuses on learning outcomes as a source of information, it may be used to correlate the contribution of a teacher or a program to the pupils' progress toward the instructional objectives. This information may then be used as research to identify correlations between different teacher (or system) behaviors and learner behaviors, or as a way of culling, improving, or encouraging continuance of certain teachers, teaching styles, or system operations.

This leads to perhaps the most politically sensitive purpose for teacher evalaution: to make a judgment concerning teacher remuneration or employment status. The ultimate goal of any teacher evaluation system is to improve instruction and increase learning. This can be done by improving teacher behavior, by clarifying needs, goals, and objectives to make instructional programs more effective and coordinated, by reallocating inputs, by positively reinforcing good teaching and negatively reinforcing bad teaching (really a part of the first), or by getting rid of the "worst" teachers according to some predetermined criteria.

III. PRESENT CONCEPTUAL FRAMEWORKS OF TEACHER EVALAUTION

Introduction

There seem to be three general trends concerning teacher evaluation--appraisal based on mutually derived objectives, appraisal based on student learnings, and appraisal based on teacher behavior. Although there is some overlap among these three approaches, they do differ in emphasis and are worthy of being viewed individually, as well as collectively.

Appraisal based on mutually derived objectives

This approach involves dialogue between a supervisor and supervisee who mutually develop goals and objectives for an approaching period of time (semester or academic year). As a consequence of this discussion, both parties agree to the identified goals and objectives, and how each is to be evaluated.

Often categories such as short range, problem solving, innovation, and personal development, are identified as means of generating objectives for each. Objectives need to be as measurable as possible, with criteria identified beforehand in order to judge how well an objective has been met. Some systems include some estimate of time and/or money needed to accomplish the stated objective.

Strengths to this approach are that it clarifies teacher and supervisor roles, removes ambiguity concerning who is responsible for what, provides a framework for continual employee appraisal with criteria spelled out beforehand, and provides an opportunity for teacher input concerning what is to be evaluated and an identification of special circumstances.

Limitations to this approach are that it can be time consuming, requires special skills by both supervisor and supervisee if it is to be successfully implemented, causes inequities to emerge between teachers concerning difficulty of and effort put into objectives, and tends to generate unrealistic objectives which can lead to frustration and a concomitant loss of morale.

Appriasal based on student learnings

This appraoch stresses that the main purpose of classroom teaching is pupil learnings. In other words, it is not what the teacher does that is so important, but rather what the student does or learns from the set of experiences provided or guided by the teachers.

Learning can be broadly or narrowly conceived. The broad perspective would include a gamut of learnings ranging from cognitive, affective, or psycho-motor to certain attitudes or values which may result indirecly from the learning experience. The more narrow view would look more at the cognitive realm of learning and place a heavy credance on the use of standardized tests.

This approach would judge the relative effectiveness of the teacher by focusing primarily on pupil outputs rather than on teacher-learner processes. Barring any extreme unethical behavior, the main concern is defining and measuring what the child has learned.

Strengths of the pupil oriented approach are that it places the emphasis on results rather than intentions, forces more careful examination of pupil needs and related learnings, ensures a higher degree of pupil involvement and forces a closer review of teacher performance.

Limitations would include that pupil learnings are difficult to quantify, more emphasis is placed on short term learnings and long range consequences are ignored, focus may be predominantly on low level cognitive skills while ignoring higher level learnings, teaching-for-the-test syndrome may result, special pupil needs and/or circumstances may not be taken into account, and unrealistic expectations may arise if normative data is used.

Appraisal based on teacher behavior

This approach places more emphasis on what the teacher and/or the learning environment does, and not necessarily on the results of such actions. An attempt is made to make more specific the desireable teacher behaviors and classroom climate.

The clinical approach, as it is sometimes referred to, would involve identifying broad areas such as planning, instruction, administration, public relations, learning environment, etc., and within each of these, spell out specifically the behaviors expected to be observed.

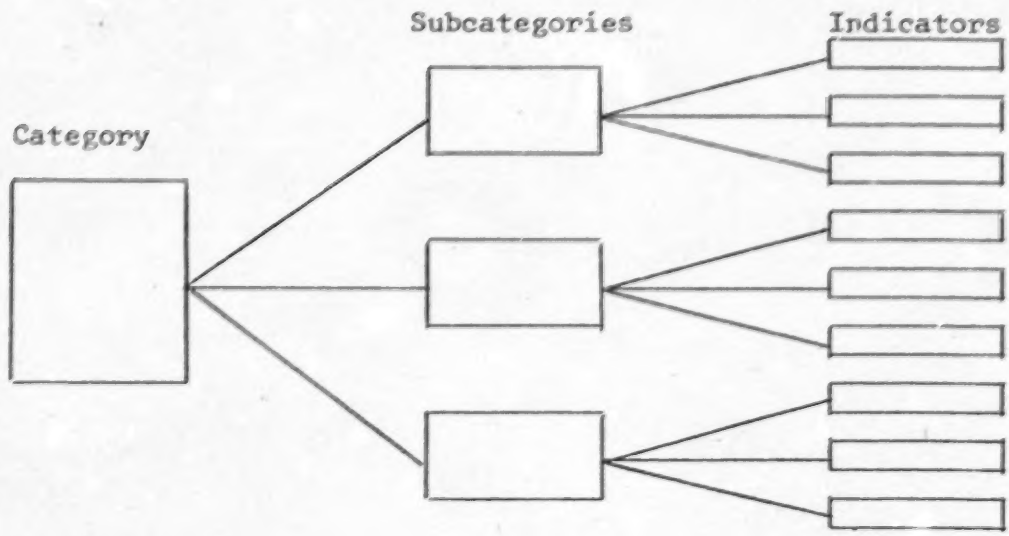
An example of this would be a category identified as instruction which could be further broken down into subcategores to include: directed toward student needs and abilities, directed to student interests (motivation), and directed toward the learning environment. In order to evaluate performance within these subcategories, specific observable teacher behavior that logically relates to the subcategory would be identified.

Again, extending the example provided thus far, the subcategory directed to student needs and abilities might include:

-the teacher provides differentiated homework assignments
-the teacher can state the strengths or weaknesses of each student
-the teacher allows for a diversity of learning styles
-the teacher offers enrichment activities.

Thus, three levels of specificity are logically developed and become a framework for establishing an overall evaluation system. Once the desired behaviors are explicated, then specific information gathering procedures can be identified.

The following graphic display illustrates the three levels of specificity of expected, or desired, teacher performance.



Strengths of this approach are that it forces a more clear definition of expected teacher performance, is based on a tighter system of logical thinking, subsequent methods of measuring teacher performance can be more systematically planned, provides a framework for comparing performance among teachers in a system, and ensures a certain equity of evaluation procedures across a system.

Limitations include: it may place an imbalanced view toward teacher behaviors versus pupil outcomes; may surface high levels of conflict in attempting to resolve what teacher behaviors should be included, may create inherent inequities when universally applied to all components of a school system, may require a high level of staff time and involvement to fully define expected teacher behavior, and generating a list of desired teacher behaviors may create unrealistic expectations.

Summary

The problem of judging performance of a classroom teacher for whatever purpose, is somewhat akin to judging a painter. Applying the three previously discussed methods for appraisal, the mutually derived objectives approach would involve the buyer sitting down with the artist beforehand, and mutually agreeing as to what the buyer is hoping the artist would paint. The buyer may be interested in a landscape versus a portrait, more oil than watercolor, and that the landscape should feature certain phenomenon as shoreline, crashing waves, large rocks, gulls, and a certain level of authenticity. The artist is free to use whatever sources, techniques, colors, etc. which he or she feels best replicates the buyer's vision.

In the appraisal based on certain outcomes, the buyer is not interested in a dialogue beforehand with the artist, nor how the artist paints, but wants a finished product that meets some criteria--such as esthetically pleasing, appropriate to home or office decor, or within a certain dollar amount. Again, it is the results of the artist's talents that counts more than how the artist paints or what the artist intended to represent in his/her work.

The last approach which focuses on behaviors and methods would examine the techniques of the artist, his/her life style, materials used, appropriate models, etc., and would judge such an artist as being knowledgeable or limited in experience concerning his/her craft.

The point being, there are multiple ways of judging a painter as well as a teacher. The more aware one is to these alternative views, the greater the possibility of knowing which appraisal process to use under what kind of conditions. Obviously, the ideal would be some combination of all three approaches, but reality may suggest some modification. Thus, the hard question becomes which approach or combination of approaches makes the most sense. Hopefully, the remaining portions of this paper will help in answering this question.

IV. FACTORS TO CONSIDER IN DESIGNING A "NEW" TEACHER EVALUATION SYSTEM

If one is considering implementing a teacher evaluation system or changing their present one, they should consider the forces which precipitated this change. Someone or some group (or groups) obviously determined what exists at present is different than what should be. The source of this perception of discrepancy between desirable and actual could be:

1. A school board who doesn't like what they think is going on in the schools
2. An administrator who perceives some low quality teaching in his district
3. A building principal who finds the students' scores on standardized tests are low and/or dropping
4. Teachers feeling a need for professional development to improve the quality of instruction and make their job more rewarding
5. Students feeling they are not getting the kind or amount of education they want from certain teachers
6. Taxpayers wanting to increase the bang they get for their buck
7. Industry complaining that recent graduates entering the job market do not have the necessary skills for entry level jobs
8. Lay persons and parents concerned about what they hear about the schools
9. Teachers feeling that the present evaluation system is outmoded and inequitable
10. Administrators feeling the present evaluation system would hold no legal clout, particularly if they try to fire a teacher with it.

At this point, an educator should make sure he is not identifying need for teacher evaluation just because a vocal minority is squeaking about some aspect of education. At such a point, the real need might be for public relations, or a community advisory council to help clarify what the public school's goals are, or diversifying the curriculum to meet the needs of varying clients. If only one group perceives a need for teacher evaluation it will probably not happen unless other groups can be convinced. Teaching is just one (though the central) component in an educational system. Another error that can be made at this point is to assume a teacher evaluation is necessary when there is a more crucial need to evaluate the whole system.

In clarifying what a school system's needs are, the purpose(s) of a "new" teacher evaluation system begins to emerge. The purposes for teacher evaluation are often vague or hidden. The best time to

ferret out any hidden agendas is at the beginning, for the clearer this intent is, the easier each step in the process of design, implementation, and evaluation of the system will be. (This is not to preclude a common and desirable occurrence of refinement, modification, or change of purpose as the system evolves.)

After one feels their initial needs and purposes for a teacher evaluation system are clearly defined, one can assess those resources within the system to support a "new" evaluation system. How much time and money can be made available? What is the level of understanding and trust between and among groups? What is the skill level of the system's personnel? Is outside help needed?

Next, one should identify what people will be involved in developing an evaluation system, and to what level. What groups should be involved in the planning process: students, teachers, department chairpersons, principals, support personnel, central administrators, board members, lay persons? Should they be elected or selected? Will they be paid, or "volunteers"? How will they be invovled: attending design meetings, responses to interviews or questionnaires, designing a part of the evaluation system on thier own (the students might design a student questionnaire as part of the evaluation system)? If the trust level is high, the time and money available is low, and the purpose of the evaluation is non-threatening, it is conceivable that one person might design, implement and evaluate the whole system. Such is usually not the case, however.

Most evaluation processes will be controversial, with an initially low level of understanding and trust among most parties. If the purposes for the initially conceived evaluation system are politically "touchy", and there is a desire for cooperation of some degree at the point of implementation, then involvement is the key to designing an evaluation system. The process should include (at least by representation) everyone upon whom the evaluation system will impact. They should be involved in the very beginning, and in every step thereafter. It may feel like a deadly long process, but it may be a necessary one. A task that might take fifteen minutes in a cooperative meeting of co-representatives might take fifteen hours if group A is there to critique the work of group B, of whom they are suspicious.

Baseline Assessment

Before getting involved in the design process, some baseline information should be collected concerning:

1. those who initiated and/or are leading the evaluation system's design, implementation, and evaluation
2. those involved in the design process
3. the system as a whole--all those who might be affected by the evaluation process

The purposes of this assessment are:

1. determine a possible starting point
2. determine present facilitators and inhibitors and thus appropriate strategies for design and implementation
3. if one is going to evaluate the evaluation system, one needs baseline data in order to measure whether there has been any change (see step 6 in discrepency model, p. 31)

The types of information that might be collected are:

1. attitudes toward teaching
2. attitudes toward learning
3. attitudes toward evaluation
4. knowledge of the system's goals, objectives and expectations of teaching behavior
5. knowledge of teacher evaluation
6. level of trust and understanding between and among groups
7. level of commitment toward changing to a "new" teacher evaluation system
8. level of skills often involved in teacher evalution: interviewing, classroom observation, item analysis
9. amount of outside training teachers are presently receiving
10. present level of student performance
11. present teacher behaviors.

Defining a Decision-Making Process

The first task of a group set up to design an evaluation system is to establish a decision making process, and perhaps other rules of communication. Examples of forms of decision making include:

a) democratic
b) consensual validation
c) negotiated agreements
d) authoritarian (edicts)

Unless the decision making process is clearly stated, there will almost surely be misunderstanding and later statements such as "But I thought we all agree to . . . ," or "Yeh, but he/she said we had to . . ."

Other rules of communication might also be implemented. If a large group is involved, input might be limited in duration (ten seconds each time you talk) or in frequency (you must give up one of your three chips each time you talk) in order to broaden the base of contributions.

Clarifying Purposes of Teacher Evaluation System

After the decision making process has been established, the purposes of a teacher evaluation system can be reaffirmed, clarified, or altered. This is the acid test of whether people are going to respect the first decision: how decisions will be made. If a superintendent agreed to a democratic decision making process and later slips in a comment, "The only limit I place upon this system is that it be implemented by September 1," or "that it allows me to discriminate between our best and our worst teachers," then there has been a breach of faith which will affect the quality of the group's output.

Identifying Alternative Approaches to Designing an Evaluation System

Once an individual or group has clarified the purpose of their evaluation sytem, it must be decided how much they will start afresh and design their own system or how much they will borrow and steal from others.

Within twenty miles of any school system there may exist at least twenty teacher evaluation forms being used to one degree or another. Why so many? Is the teaching/learning style at various schools and levels so varied as to warrant this? A search of many teacher evaluaion instruments shows a striking consistency in both form and content. Why all the forms then? Three possibilities come to mind:

1) Ignorance of the existence of other designs
2) Need of professional evaluators and central office staff to stay employed by continuing to help individuals, schools, or districts to (re-)design instruments
3) A need for feelings of ownership by practitioners.

While all these possibilities are somewhat regrettable, we will dismiss the first two as being just that, and concentrate on the third--ownership--with which anyone involved in introducing an evaluation component must contend.

One advantage to employing an evaluation system that has already been designed, is efficiency--a saving of time and money. The quality of the system in terms of logical consistency and measurability of indicators may also be higher using a borrowed rather than a built-from-scratch system.

The argument for content validity can go either way. A system designed by others may have more validity in terms of allowing for a diversity of teaching styles or being somewhat connected to educational research. However, a locally designed system may have more validity because it better reflects the values, goals, and objectives of those using it. This is especially true if different evaluation instruments are used for different instructional assignments. Our

experience, however, has been that most teacher evaluation instruments are very similar in content.

The main disadvantage of borrowing someone else's instruments is the ownership issue. Before people can feel comfortable using an evaluation instrument, they must feel it is their's. Their pride and identification with having built it reduces the threat level.

Therefore, the crux of the issue on approaches to designing an evaluation system seems to be between efficiency and a feeling of ownership. There are many middle grounds between building a system from scratch and adopting someone else's lock, stock, and barrell. One alternative might be to have someone with interest, knowledge and skills in teacher evaluation generate a list of high quality items and indicators, and have a group select from that list. We all know that getting people to state their objectives in measurable language, or design good survey questions, can be a long process. If someone else can do this homework, the efficiency is gained without losing those people's involvement.

What Type of Data Needs to be Collected

The purpose for a teacher evaluation system will determine the type of data which will be measured and recorded. If some foresight is not given to this issue, the converse will be true: the way in which observations have been recorded or measured will determine the purposes for which the system can be used.

Data is usually characterized into four types, depending on the level of measurement employed to collect the data: nominal, ordinal, interval, and ratio. Nominal data is merely a classification: yes/no; present/absent; center/forward/guard. It tells whether an attribute, or group of attributes, is possessed or not. Ordinal data indicates the relative quantity of an attribute on some scale: poor/fair/satisfactory/good/excellent; always/sometimes/never. While it does indicate possessing more or less of some attribute, it does not imply equal intervals between each point on the scale. The difference between fair and poor is not assumed to be the same as the difference between excellent and good. Interval data measures the amount of an attribute such that the increments between points are equal. Ratio measurements have an absolute zero point along with an equal interval scale. This type of data is rarely used in education. Examples of ratio data, however, might include measurement of noise in a classroom using decibals, or measuring the distance a youngster can throw a softball.

Making educational assessments may require only nominal data (i.e. determining the number of elementary teachers who use math workbooks as opposed to those who don't). Or one might want to measure how many American history teachers state their primary criteria in grading students as being standardized tests, teacher designed tests, homework performance, or classroom participation. Assessment may also involve ordinal, interval, or ratio data.

If evaluation is to be a rational process it must at some point deal with either actual or implied ordinal data. One can collect nominal data, but if one makes a judgment (evaluation) concerning this data, it acquires a value (often not clearly stated). For example, an evaluator could classify teachers as to whether or not they are presently carrying on any action research in their classroom. This is nominal data. But if there is a judgment involved in this observation (i.e., teachers who carry on action research are more professional or better instructors than those who don't), it is really ordinal data and could just as easily be written as:

Amount of professional development as indicated by the presence of action research in the classroom:

0 10

none very much

Any scale values could be assigned, but they would be ordinal.

Most evaluation is followed by decisions, and these decisions often require ranked information:

1) which in-service need is the greatest?
2) what are my weakest qualities as a teacher?
3) what are the department's primary instructional objectives?
4) who are our best second grade teachers?

There is often great resistance to rating teachers with numbers. In evaluating a teacher concerning the indicator "employs a variety of learning materials," people feel much more comfortable checking poor/fair/good/excellent, than 1/2/3/4. Though the initial behavior of quantifying subjective judgments may be difficult, it may make subsequent (and more significant) decisions considerably easier to make. This is assuming people want to make those decisions; the argument against quantifying teacher evaluation is a great one for avoiding making difficult communications and decisions because you don't have the necessary information.

A note of caution here. Though determining salary or employment status may require ranked information, the addition or subtraction of ordinal numbers is not permissible in statistics. While satistically it is a no-no to compute ordinal data, all parties involved may agree that it is more satisfactory than having a superintendent eyeball the results of a twenty-item student questionnaire in order to determine who his least competent middle school science teacher is.

Sources of Information

Theoretically, all persons who have an opportunity to directly observe the performance of a classroom teacher should be solicited for perceptions about such performance. Pragmatically, however, this isn't always plausible, and is fraught with potential fears and misunderstandings about persons' motivations in reporting their observations.

It should be kept in mind that the best evaluation of any individual's performance is the balanced evaluation which draws upon numerous samples of behavior and numerous sources of information. To rely on one person making only a few observations over a long period of time can only provide a very limited view of a person's performance, regardless of how diligent and sincere the motives of such a single observer. Thus, it is urged to consider multiple sources of information which include any of the following: supervisors, other teachers, teacher-self, pupils, support service personnel, parents and miscellaneous persons (such as consultants who may have opportunity to relate first hand with a given teacher's activities). It must be emphasized that sources of information relate directly to the specific behavioral indicators and should be selected for their logical relationship.

1. Supervisor

This source may include principals and department chairpersons who have sufficient time to observe the teacher in a variety of contexts, both in and out of the classroom. Supervisors are a valuable source of information--particularly if they have sufficient expertise in areas related to the teacher's role, are sufficiently informed through direct observation and interaction with the teacher, and have received some training in the conduct of a supervisory role.

A caution to be considered is that supervisors should not be the only source of information or considered automatically as the best source of information.

2. Other teachers

Often overlooked for a variety of reasons are peers. These persons have daily contact with one another in various situations and have valuable insights which often go untapped (because of certain professional taboos?). This is an unfortunate circumstance. Techniques can be explored which can ensure some control over the selection of other teachers for feedback purposes and which ensures a degree of anonymity if desired. Colleagues can provide valuable feedback if a mechanism can be found to gain such data.

One plausible approach might be for a teacher to nominate potential persons best suited to judge a teacher's performance, and the supervisor pick a predetermined number of persons at random from the proposed list of names.

3. Teacher-Self

Often overlooked is the teacher him/her self. Personal interviews are considered essential in the hiring of personnel, but a similar effort is often not made thereafter to illicite from a teacher information which probably cannot be gained from any other person.

It is a matter of knowing in advance what information is desired, and soliciting data from the best sources, including the teacher being judged.

4. Pupils

A great deal of controversy surrounds the use of pupils to assess the performance of teachers. There is increasing evidence students are capable of making fair and informed judgments of a teacher's performance. Certainly much depends on what and how students are asked to evaluate a teacher. If students are asked questions which are clear, and upon which they are sufficiently informed, then it seems essential to consider what they can contribute to the overall evaluation.

The important consideration to keep in mind while identifying sources of information is a close logical linkage should exist between the teacher-behavior (deemed necessary to evaluate) and the best source(s) of information. For example, if the evaluation system places a heavy emphasis on positive teacher-parent relations, it seems necessary that parents be considered a primary source in judging how well a teacher performs his/her responsibilities in this area. This is not to say other individuals who may have an opportunity to observe the teacher in interaction with parents should be overlooked. When Sears, Roebuck & Company attempts to evaluate the effectiveness of their service repairmen in customer relations, they rely on direct feedback from the customer, not the supervisor who may never observe their servicemen in direct contact with customers.

5. Support Service Personnel, Parents, and Miscellaneous Persons

As often is the case, many wise and prudent persons come in daily contact with classroom teachers who may have valuable perceptions which go untapped. From a logical and rational point-of-view, no primary source of information should be overlooked.

Again, certain procedures, as suggested under "other teachers" might offer sufficient control and confidence in using these sources.

Data Collection Methods

As with sources of information, once the desired teacher behavior or results have been identified, then logical methods for becoming informed about these behaviors can be explored.

Below are listed a variety of methods for gathering information along with strengths and weaknesses of each. Which of these approaches is to be used will depend somewhat on local situations in terms of clerical personnel, training of persons in their use, time and money available, to name just a few.

1. Data Collected by a Mechanical Device (e.g., audio or video tape)
   Advantages:
   1. Avoid human errors.
   2. Stay on job - avoid fatigue.
   3. May capture content missed by written records (e.g., voice inflection).

   Disadvantages:
   1. Cost.
   2. Cannot make independent judgment.
   3. Complexity can cause problems in operating devices.

2. Data Collected by an Independent Observer
   Advantages:
   1. Can be used in natural or experimental settings.
   2. Most direct measure of behavior.
   3. Experienced, trained, or perceptive observers can pick up subtle occurrences or interactions sometimes not available by other techniques.

   Disadvantages:
   1. Observer's presence may cause an artificial situation.
   2. Hostility to being observed.
   3. Inadequate sampling of observed events.
   4. Ambiguities in recording.
   5. Frequent observer unreliability.

3. Data Collected by Written Accounts
   Advantages:
   1. Can use critical incident technique, eliminating much "chaff."

   Disadvantages:
   1. Hard to be complete.
   2. Hard to avoid writing interpretation as factual data (e.g., "Mary kicked John because she was angry with him.")

4. Data Collected by Observation forms (e.g., observation schedules)
   Advantages:
   1. Easy to complete; saves time.
   2. Can be objectively scored.
   3. Standardizes observations.

   Disadvantages:
   1. Not as flexible as written accounts - may lump unlike acts together.
   2. Criteria for ratings are often unspecified.
   3. May overlook meaningful behavior that is not reflected in instrument.

5. Data Produced by the Subject Himself: Self Reports
   Advantages:
   1. Can collect data too costly otherwise (e.g., eliminates

the endless observation necessary to really get to know a person's philosophy, attitudes, etc.).
2. Can collect data not accessible by any other means (private thoughts, feelings, actions, emotion-laden material).

Disadvantages:
1. Depends on respondent's awareness of self.
2. Depends on respondent's honesty and/or security.
3. Depends on respondent's "accurate memory" when dealing with past events (selective recall).
4. May necessitate anonymous responses where threat is perceived.

6. Data Produced by Interviews

Advantages:
1. Interviews are adaptable to a wide variety of respondents, topics, and situations.
2. Interviews are uniquely suited to in-depth exploration of an issue. The unstructured interview is informal and allows the interviewer to pursue interesting answers and to probe below the surface.
3. The structured interview, which provides a detailed guide to topics and often required wording and sequence of questions, can be used when information from various interviews must be as comparable as possible.

Disadvantages:
1. Each interviewer and each interview is a little different, and there is no completely practical way to control or estimate the effect of these differences.
2. Interviews require a lot of time, energy, and money. Thus, data is usually collected only from a small number of people.
3. Interviews require a very high degree of skill in planning and execution.
4. Interview data is often difficult to summarize and interpret.

7. Data Produced by Questionnaires

Advantages:
1. They are an economical way of gathering a large amount of data.
2. Data can be collected by mail.
3. They are particularly well adapted to sampling techniques. The sampling plan, and not limitations of the process itself, is the prime factor in the sampling decision.
4. Anonymity is possible and encourages honesty and frankness in answering.

Disadvantages:
1. Unbiased or neutral phrasing of items is difficult to achieve, particularly in controversial areas.
2. People are seldom equally well informed about the problem.
3. Questions must be kept simple, which limits the quality of the information obtained.
4. The longer the questionnaire, the lower the return rate.

4. continued....
   The shorter the questionnaire, the smaller the amount of information. The compromise is always difficult.
5. Valid generalizations cannot be made unless a high rate of return is obtained.

8. Data Produced by Rating Scales and Check Lists
   Advantages:
   1. They are particularly well-adapted to improving the validity and reliability of on-site observation of actual behavior.
   2. They are easily reused, and thus provide data for interpreting change.
   3. They can be used in group situations.
   Disadvantages:
   1. They are limited to behavior, and are difficult to use when there is interest in attitudes or achievement.
   2. There is a tendency to avoid extreme ratings.
   3. The data is affected by the conscientiousness, severity, experience, and physical state of the rater. This will be different both among raters and for a single rater over time.
   4. The description of what is to be rated is often vague.

9. Data Produced by Unobtrusive Measures: Records
   Advantages:
   1. Records are permanent and usually fairly well up to date.
   2. The only cost of collection is clerical.
   3. They are readily accessible (assuming no legal problems).
   Disadvantages:
   1. They are appropriate only to a limited number of objectives.
   2. There is usually a lot more information than can be used, which requires an element of selectivity.

10. Data Produced by Unobtrusive Measures: Unobtrusive Observation
    Advantages:
    1. They are particularly valuable for obtaining data about attitudes.
    2. They are appropriate in group situations.
    3. They avoid stimulating students, etc., to work harder because they know they are being observed.
    4. See also comments on rating scales and check lists.
    Disadvantages:
    1. In some circumstances, the method is akin to spying; and offensive to some.
    2. See comments on rating scales and check lists.

11. Data Produced by Unobtrusive Measures: Accretion and Erosion
    Advantages:
    1. They are particularly valuable for obtaining data about attitudes unbiased by student feelings that they are being measured.
    2. They are generally very inexpensive.

Disadvantages:

1. They require more imagination to devise than most measures.
2. They are vulnerable to sampling biases.

12. Data Produced by Evaluation Committee

Advantages:

1. The committee can draw on the expertise of several people.
2. Individual biases are usually eliminated when challenged by someone else on the committee and the position cannot be defended.

Disadvantages:

1. It is generally expensive in terms of time.
2. The committee is often monopolized by one or two vocal members.
3. The approach is often random and non-systematic, the results disorganized and difficult to use.

13. Data Produced by Community Groups

Many community groups have important opinions and valuable information. This information can often be collected through attendance at meetings, copies of minutes, publications, reports, interviews with officers and the like.

Advantages:

1. Such groups are usually sincere and genuinely involved in, and concerned with, school and community problems.

Disadvantages:

1. The groups often exist to prove a point of view. Their position requires a great deal of confirmation before it is to be believed.
2. Their evidence is usually anecdotal, and thus could be only the spectacular exceptions to the norm.

Management Structures

Depending again upon the purposes of the evaluation system and levels of interest and trust, various management structures may evolve. So far, the most common model has been hierarchical: principals or department chairperson evaluating teachers under their supervision.

Other models may be conceived:

1. An individual teacher could perform a self-evaluation. That person could manage their own information system or use outside consultants to come in to provide feedback.
2. Evaluation could be managed by peers (such as a teacher organization) interested in professional growth.
3. An educational system could hire or contract an individual or group whose specific responsibility would be staff evaluation.
4. A group of parents cooperatively running a private school might take charge of evaluating their staff.

Equity vs. Diversity

At the same time as one is deciding upon possible sources of information, it must be decided how many evaluation systems must be designed to fulfill the purposes for teacher evaluation in that system.

Efficiency says that the fewer evaluation systems, forms, or sources of information, the better. Also, if employment and/or salary status are the purposes of teacher evaluation (and, thus, teachers must be ranked), equity demands that all teachers being compared concerning professional competence should be evaluated using the same criterion and instruments.

However, there usually exists a wide diversity of grade levels, programs, and teaching and learning styles within a system, and most educators tend to value and encourage that diversity. A standardized evaluation format has the potential to direct and narrow this diversity. Also, the more generalizable an evaluation instrument is, the less valid it may be for any specific program, teacher, or learner.

A system with twenty professional positions and twenty evaluation procedures may be wasting time in both the design and the implementation of staff evaluation. Yet it is difficult to design a form that is specific enough to have any validity or ability to discriminate between high and low quality performance if it must be applied to K-12 classroom teachers, nurses, music teachers, etc.

Selecting Plausible Design Features

The person or group that has ariticulated the alternative to a teacher evaluation design must then decide upon the most plausible approach for their particular purposes and resources. Are they going to build a teacher evaluation system from scratch, model certain components after pre-existing systems, or borrow a design outright from someplace else? Are they going to need nominal, ordinal, or iterval data on which to base their evaluation judgments? Will there be one evaluation format for all professionals being evaluated, or will there be several? From where will the data for evaluating teacher effectiveness come?

Rarely will these decisions be "clean" ones. There will be trade-offs lost with the selection of any course of action. This frustrating reality, however, should not discourage some action. It is very easy for an individual or group to decide that because of these ambiguities, the idea of designing a "new" system should be abandoned "for now." The known but often unsatisfactory present practices can look awfully appealing when faced with an unknown future practice. Where possible, leadership should be taken to point out that without taking some risks, no significant change can

take place. Here it may also be worth reinforcing the fact that the "new" evaluation system should itself be evaluated upon implementation in order to make rational changes in its design or process.

"This will take too much time" is an often heard statement in evaluation design meetings. Whether this is an alibi, an expression of anxiety concerning an unfamiliar and seemingly difficult task, or an accurate assessment should be ferreted out.

More accurate estimates of time involved can be obtained by a task analysis on paper or an actual pilot test. A planning environment that encourages people to express their apprehensions with the new can reduce the need to create rationales as concerns.

Individuals or a whole system can be analyzed as follows:

1. What percentage of time do you presently spend on evaluation?

2. Given the priority you feel concerning staff evaluation, how much time should you (would you be willing) to spend on it?

3. What are you presently doing that you feel is less important than staff evaluation?

Building the Evaluation System

Whether the new teacher evaluation system is built or borrowed, if the observation of teacher or pupil behavior is chosen as a source of information for evaluating teachers, several more decisions must be considered.

Specificity (Reliability) vs. Information Overload

Generally speaking, the more specific the behaviors are which are used as indicators of desirable behavior, the more reliable the evaluation system will be. That is, the more simple and discrete a behavior is, the more likely two people (or one person from one day to the next) are to agree that that behavior is absent or present. Even this fact, however, is open to some debate.

To rate a teacher on "good student/teacher relationships" would illicit diverse ratings depending on the values and mood of the evaluator. There would probably be more agreement among evaluators on "evidence of many pupils participating in class." Still more reliability might be achieved with the indicator "teacher supplies positive reinforcement to students who constructively contribute in group activities." There might be near total agreement on an item such as "after a student has offered a fact, feeling, or concept that

has not been previously verbalized during class, the teacher will obtain eye contact with the student, move his head in a vertical axis, and say 'good'." We can see that though the indicators became more behaviorally specific and measurable, and thus more reliable, it becomes questionable whether the final behavior has much to do with the all-be-it vague goal of "good student/teacher relationships."

Ideal vs. Practical

As more sources of information are employed to allow for equity (fairness) and validity, and as indicators of good instruction are made more discrete and measurable to assure reliability, a new danger arises. While equity, validity, and reliability are valid goals, they may lead to a state of information overload which makes the evaluation system so long or obtrusive that usage is discouraged. If in getting to the ideal, one passes by the practical, one still has a way to go. One approach might be to design an ideal system and then edit it back to practicality. Another would be to set initial reality constraints to the design product and to build within them (allowing for redefinition of constraints as the process unfolds).

Weighing Information

A question may arise as to whether different data sources for teacher evaluation are equally important or valid. An appropriate response to this might be to weigh the overall effect of scores on one instrument (say parent interviews) as against another (say classroom observation).

A very important issue also arises concerning differences between evaluators in both severity and differing values (if the judgments are subjective). A certain subjectivity cannot be bleached out of any teacher evaluation system, but some differences between evaluators can be compensated for. One approach might be in-service training for evaluators and/or refinement of the instrument until a high degree of inter-evaluator reliability is obtained. Another solution might be to accept those differences, "normalize" them, and make any evaluations relative to the evaluators' norm. Take two evaluators, A and B, using the same classroom observation instrument to rate teachers on a scale from -50 to +50. A's average rating of a teacher is -5. B's average rating of a teacher is +30. A teacher rated by evaluator A as "0" may indeed be a better teacher than another rated by B as "20". Using the difference from this evaluator's mean as an "adjusted" score is a solution, though it is assuming that ordinal data is interval data. While this is not true, as long as all parties agree to some rules for weighing and comparing scores and do not make false assumptions about their results, no sacred rules have been disobeyed.

V. IMPLEMENTING TEACHER EVALUATION SYSTEMS

Introduction

Obviously there are many roads to Rome. That is, there are different ways of achieving similar ends. Presented in Section V are a series of stages which offer at least one way to achieve the implementation of a teacher evaluation system.

The need to involve persons who will either be responsible for conducting evaluations, or will be the target of evaluation, is stating the obvious. What is attempted here is to spell out more specifically in what areas and what levels such involvement seems desirable. It is possible to view this involvement in three stages: preparation stage, implementation stage, and summary stage.

Preparation Stage

In Section IV, is spelled out the various activities which call for involvement of different persons in the design of any evaluation system. It seems wise to suggest a stage, which follows the design stage and precedes the implementation stage, which may be referred to as the preparation stage.

Effectiveness of any system designed is contingent upon adequate preparation of different personnel to carry out their responsibilities. Thus, this stage has the primary focus of training persons for their respective roles in executing the designed teacher evaluation system.

This training might cover two broad areas referred to as skills and knowledge. Included in skill acquisition are such general skills as interviewing, observation, questionnaire construction, synthesizing data, and drawing inferences. In a more specific sense, each person should become trained in the use and application of the newly designed system. This should involve more than a one hour faculty meeting.

In the knowledge area might be included such items as alternative data collection procedures and instruments, sampling techniques, sample teacher evaluation systems, related research results, and issues related to teacher evaluation.

This preparation stage should increase everyone's awareness to the dynamics of teacher evaluation and provide adequate knowledge to understand the rationale behind various methods being followed.

Implementation Stage

To maximize communications among all parties and to maintain a high level of trust, the following steps are suggested as one model:

Step 1 - Explanation of Evaluation System to staff
Re: Format
Role of teacher
Role of supervisor, if involved

Step 2 - Development of Evaluation schedule to include Preconference, data collection, and post conference.

Step 3 - Implement Preconference Sessions
Re: Item clarification
Information gathering procedures

Step 4 - Implement Observation and Information Gathering Procedures.
Re: Visitations
Student Survey
Interviews - teacher(s), administrators, others
Student testing

Step 5 - Implement Post-Conference Sessions
Re: Results
Interpretation of data
Recommendations and commendations
Future development activities

Step 6 - Submit Written Report by Supervisor to principal and/or superintendent
- teacher sign off to acknowledge content of report

Step 7 - Submit Written Rebuttal
Re: negative data, by teacher to superintendent (optional)

Summary Stage

This stage needs to address those who are responsible for gathering, synthesizing, and making data about the teacher available. In other words, how much responsibility does the classroom teacher have in pulling together data from varying sources, and how much responsibility does the supervisor have.

If two parties were involved in the evaluation system, and the trust level was not optional, it might become necessary to identify some person, other than the teacher, through whom forms, questionnaires, etc., are processed and filed. It is assumed all interested parties, including the teacher, would have access to such files. Depending on the resources of a school system, either the principal, or his/her designee, would have such responsibilities. It seems important that the design include procedures for gathering the information and summarizing for reporting purposes.

VI. EVALUATING THE TEACHER EVALUATION SYSTEM

Introduction

Regardless of what system is ultimately designed and implemented, a process by which it can be evaluated needs to be thought of prior to implementation. There are many dimensions which any evaluation approach might include, and each of these should be carefully considered. Basically, the meta evaluation should attempt to demonstrate the merits or demerits of the new system.

Resources of time, money, and energy will either facilitate or delimit the magnitude of the meta evaluation. Proposed in this section is a comprehensive model--parts of which may be excluded, depending on local circumstance.

The model proposed is a discrepancy model which attempts to examine the inputs, processes, and outputs of the teacher evaluation system. An attempt is made through this model to ascertain the effectiveness and efficiency of the newly designed teacher evaluation system in meeting the purposes for which it has been designed. This approach calls for a high level of planning prior to implementation.

The meta evaluation model proposed here attempts to place high value in a formative type of evaluation, as opposed to a summative type of evaluation. The meta evaluation should permit mid-course corrections while the new system is being tried out, rather than waiting to the very end.

Discrepancy Model

The following paradigm is used as a framework for organizing a meta evaluation of the selected teacher evaluation system. Basically it provides a focus on the discrepancy, if any, between desires inputs, processes, or outputs and actual performance.

	Desired	Actual
Input	②	④
Process	③	⑤
Output	①	⑥

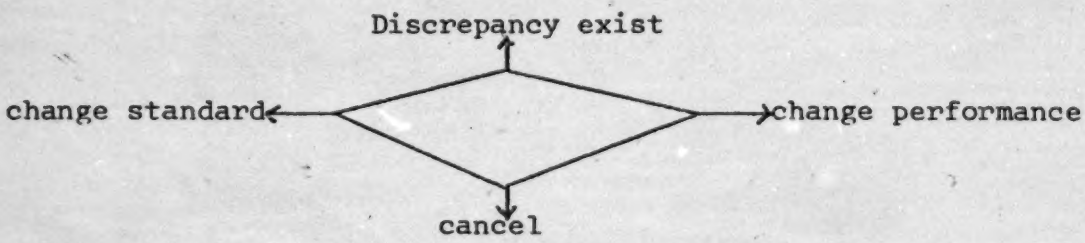
Step 1 - Define desired outputs of "new" system
- A. Purpose(s)
- B. Attitudes
- C. Reports
- D. Others

Step 2 - Define desired inputs of "new" system
- A. Personnel needed - teachers, administrators, secretarial
- B. Time needed - meetings
- C. Money needed
- D. Special equipment
- E. Others

Step 3 - Define desired processes of "new" system
- A. Activities to be conducted
- B. Methods used
- C. Data Collection procedures
- D. Others

Step 4 - Assess actual inputs of new system

At the onset of implementing the new system, an attempt should be made to assess what inputs were actually provided. Were the funds, in-service meetings, equipment, etc., provided as was anticipated? If a discrepancy is evident between desired and actual inputs, the following decision-making model is suggested.



This model suggests there are three choices: change the standard, change the performance, or cancel. If sufficient funds were allocated for the new teacher evaluation system, for example, then either an attempt should be made to elicit the desired funds, or to modify performance expectations in light of the funds allocated, or it may be judged without sufficient funds which are not likely to be acquired that the new system should be scratched.

Step 5 - Assess actual process of new system

At some mid-point in the implementation of the new system, an assessment should be made to determine if the processes or methods previously identified are being followed or not.

Again, if a discrepancy is observed between actual and desired performance, the decision-making model discussed in step four would be followed.

Step 6 - Assess actual outputs of "new" system

Near the end of a reasonable trial period, two or three years, an output evaluation should be conducted. This would involve returning to the data developed in step one, desired outputs, and determining how well these outputs were achieved, comparing them with baseline data generated when initially designing the evaluation system.

If base line data were collected on such aspects as attitudes toward evaluation, student performance on standardized tests, etc., then the data collected in this stage can show what change, if any, occurred over time.

Step six should also allow for the identification of unintended consequences. In other words, there were certain intended outcomes identified in step one, but as with most projects involving humans, often there are other results which were not anticipated and yet were significant in the lives of those involved. Outcome evaluation should allow for this to be examined.

Again, the data from step six should be subject to the same discrepancy analysis as in steps four and five before it is determined to continue the new system, or modify it, or cancel it and start all over. Each of these are reasonable choices and whatever meta evaluation is conducted should permit clarity in making such a choice.

VII. RELATED READINGS

Allen, Paul M., et al., Teacher Self Appraisal: A Way of Looking Over Your Own Shoulder, Worthington, Ohio: Charles A. Jones, 1970.

Amidon, Edmund J., and Flanders, Ned A., The Role of the Teacher in the Classroom: A Manual for Understanding and Improving Teacher Classroom Behavior, Minneapolis: Association for Productive Teaching, 5408 Chicago Avenue, So., 1971.

Beegle, Charles and Brandt, Richard (Editors). Observational Methods in the Classroom. Washington, D.C.: A.S.C.D., 1973.

Blumberg, Arthur, Supervisors and Teachers: A Private Cold War, Berkeley: McCutchan, 1974.

Bolton, Dale L., Selection and Evaluation of Teachers, Berkeley: McCutchan, 1974.

Bridges, Edwin M. "Faculty Evaluation--A Critique and a Proposal." Administrators Notebook, Vo. XXII, No. 6, University of Chicago, 1974.

Cartwright, Carol and G. Phillys Cartwright. Developing Observation Skills. New York, New York: McGraw-Hill Book Co., 1974.

Cogan, Morris. Clinical Supervision. Boston: Houghton Mifflin Co., 1973.

Fox, Robert, Luszki, Margaret, and Schmuck, Richard. Diagnosing Classroom Learning Environments. Chicago: Science Research Associates, Inc., 1966.

Gallagher, James, Nuthall, Graham, and Rosenshine, Barak. Classroon Observation, AERA Monagraph Series on Curriculum Evaluation. Chicago: Rand McNally and Co., 1970.

Glass, Gene V. "A Paradox about Excellence of Schools and the People in Them", Educational Researcher. American Administration Educational Research Association, March, 1975.

Goldhammer, Robert, Clinical Supervision, New York: Holt, Rinehart and Winston, 1969.

Gorman, Alfred H. Teachers and Learners: The Interactive Process of Education. Boston: Allyn and Bacon, Inc., 1969.

Griffith, Francis, A Handbook for the Observation of Teaching and Learning, Midland, Michigan: Pendell Publishing Co., 1973.

Lewis, James, Jr. Appraising Teacher Performance. West Nyack, N.Y.: Parker Publishing Co., Inc., 1973. (An "evaluation and development by objectives" approach)

Magner, Robert, F., and Pipe, Peter. Analyzing Performance Problems or You Really Oughta Wanna. Belmont, California: Fearon Publishers, 1970.

McNeil, John D. Toward Accountable Teachers: Their Appraisal and Improvement. New York: Holt, Rinehart, and Winston, Inc., 1971.

Mosher, Ralph L., and Purpel, David E., Supervision: The Reluctant Profession, Boston: Houghton Mifflin, 1972.

National School Public Relations Association, Evaluating Teachers for Professional Growth, Arlington, VA: 1974.

Popham, W. James. Designing Teacher Evaluation Systems. Los Angeles, CA: Instruction Objectives Exchange, December, 1971.

Popham, W. James. Evaluating Instruction. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

Redfern, George, How to Evaluate Teaching: A Performance Objectives Approach, Westerville, Ohio: School Management Institute, 1972.

Sarthory, Joseph A. Professional Improvement and Staff Evaluation. Kansas State Department of Education, 120 East 10th Street, Topeka, Kansas.

Shaw, Jane S. "Students Evaluate Teachers and It Works", Nation's Schools, Vol. 91, No. 4, April, 1973.

Simon, Anita and E. Gil Boyer. Mirrors for Behavior III, An Anthology of Observation Instruments. Communications Materials Center, Inc., 1974.

Walberg, Herbert J. (Editor) Evaluating Educational Performance. Berkeley, California: McCutchan Publishing Corporation, 1974.